

Optimisation débit-distorsion d'un schéma de codage vidéo orienté-régions

Stéphane PATEUX¹, Claude LABIT¹

¹IRISA/INRIA

Campus universitaire de Beaulieu, 35042 Rennes, France

{spateux,labit}@irisa.fr

Résumé – Dans cet article, nous présentons une technique d'optimisation d'un schéma de codage orienté-régions pour le codage de séquences vidéo. Cette optimisation est effectuée via un étiquetage pour chaque image d'une segmentation spatiale de celle-ci ; les différentes opérations réalisant l'étiquetage, ainsi que les variations de pas de quantification sont contrôlées via un critère de type débit distorsion afin d'obtenir une solution optimale. Le schéma de codage ainsi développé s'avère être assez performant, notamment comparé aux approches existantes (MPEG1,2 H26x).

Abstract – In this paper, is presented an optimization technique for region-based video coding scheme. This is perform by a region labelling of an over-segmentation of the various images; labelling steps as well as quantization level search are sorted according to a rate-distorsion criterion in order to get an optimal solution. The video coding scheme thus obtained is quite efficient, especially when compared with existing coding schemes such as MPEG1,2 and H261,3.

1 Introduction

Dans le cadre des schémas de codage orienté-régions, le point délicat est la définition du partitionnement en régions utilisé pour la compensation en mouvement dans la phase de prédiction. Les premières études sur ce sujet proposaient d'utiliser des techniques de segmentation spatiale [7], basée mouvement [3], ou bien encore spatio-temporelle [2]. Cependant ces techniques ne prenaient pas en compte l'aspect débit-distorsion, et donc ne donnaient pas toujours des résultats intéressants en terme d'efficacité de compression.

Des études plus récentes([5, 6]) se sont alors intéressées à une optimisation d'un point de vue débit-distorsion d'algorithmes de segmentation utilisés dans des schémas de codage orienté-régions, en recherchant le compromis entre une segmentation grossière (faible coût de codage, mais prédiction pauvre), et une segmentation fine (coût de codage élevé, bonne prédiction). Ces études portaient sur des segmentations de type quadtree ou à base d'arbres de partitions pré-définies (hiérarchie de segmentations spatiales emboîtées). Dans ces études, l'optimisation débit-distorsion est effectuée en fusionnant au mieux les régions (les fusions possibles étant limitées du fait de l'existence d'un arbre de partitions).

Dans [4], nous avons proposé une autre technique d'optimisation dans le cas d'un codage à qualité constante, basé sur le formalisme MDL, non contrainte par la donnée initiale d'une hiérarchie de partition. Nous proposons alors dans cet article d'étendre au formalisme débit-distorsion cette précédente étude, et de pouvoir proposer des segmentations optimisées pour des schémas de codage orienté-régions fonctionnant à qualité ou à débit constant.

2 Optimisation débit-distorsion de la segmentation

Dans [1, 4], afin d'optimiser la segmentation, une technique d'étiquetage de régions sur une sur-segmentation spatiale initiale était utilisée. Les régions étaient regroupées en macro-régions afin de diminuer le coût de codage global. Pour se faire, des fusions de régions ainsi que des relaxations d'étiquettes étaient utilisées. Toutes ces opérations étaient ordonnancées selon un critère de type débit-distorsion; i.e. à chaque opération i étaient associées une variation en coût de codage ΔDL_i et une variation de l'erreur de reconstruction ΔEQM_i , l'opération minimisant le rapport $\frac{\Delta EQM_i}{|\Delta DL_i|}$, sous la contrainte d'une diminution de débit, était alors retenue. Cette technique peut être comparée à l'approche classique d'optimisation débit-distorsion à base de multiplicateurs de Lagrange. En effet, dans ce dernier cas, on cherche à minimiser une fonctionnelle du type $D + \lambda R$, où D représente la distorsion (EQM) et R le débit. Si l'on procède alors à une recherche de l'optimum par augmentation progressive de λ , les premières opérations de fusions ou relaxations seront telles que: $\frac{\Delta EQM_i}{|\Delta DL_i|} < \lambda$; les premières opérations auront donc un critère $\frac{\Delta EQM_i}{|\Delta DL_i|}$ inférieur aux suivantes.

Par ailleurs, lors de la réalisation d'une fusion de régions, le mouvement associé à la nouvelle région créée peut être choisi au mieux parmi plusieurs mouvement disponibles (mouvement de l'une des deux régions, combinaison des deux mouvement, ou bien encore version raffinée du mouvement) au sens du critère débit-distorsion le plus favorable. Cette technique permet ainsi d'obtenir de meilleurs mouvements pour la prédiction car choisis pour leur performance en terme de codage, et non en terme de

qualité de prédiction.

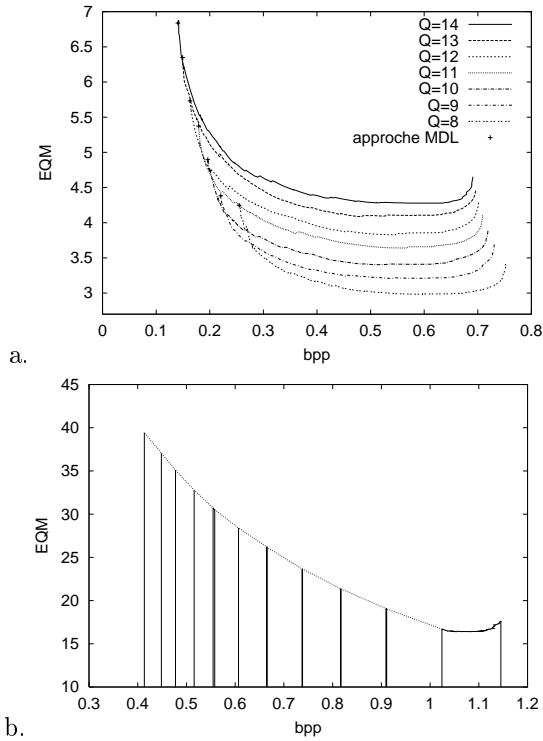


FIG. 1: Évolution du débit lors de la recherche de la segmentation optimale. (a) pas de quantification fixe (séquence Suzie), (b) avec variation du pas de quantification (image 24 de la séquence Foreman), les traits verticaux délimitent les zones où le pas de quantification ne varie pas.

La figure 1.a présente l'évolution typique du débit et de la distorsion lors de la phase d'étiquetage pour différents niveaux de reconstruction (évolution des hauts débits vers les faibles débits). Comme les opérations de fusion et de relaxation sont effectuées tant qu'il y a diminution de débit, on peut voir que cette technique n'est pas forcément optimale, et qu'il aurait été intéressant de pouvoir changer de pas de quantification.

Aussi introduisons-nous alors la possibilité de pouvoir incrémenter le pas de quantification utilisé dans la phase de codage de l'erreur de prédiction, afin de rechercher une solution plus proche de l'optimale. Cette nouvelle opération est ajoutée aux deux autres types d'opérations, et mise en concurrence avec ces dernières dans le même ordonnancement débit-distorsion.

Afin d'arrêter la phase de recherche de la segmentation optimale, un objectif en débit, ou en qualité est donné, et l'algorithme s'arrête dès que l'objectif est atteint. Cet objectif est mis à jour d'image en image afin d'assurer une régulation à l'aide d'une simple loi de commande proportionnelle: $T_{t+1} = T_t + (T^* - O_t)$ où T_t est l'objectif fixé à l'instant t , T^* , l'objectif global à atteindre et O_t , le résultat obtenu à l'instant t .

La figure 1.b montre l'évolution typique du débit et de la distorsion pour cette version modifiée de l'algorithme. Après une première phase consistant principalement en

des fusions et relaxations d'étiquette de régions, on n'observe presque plus que des ajustements du pas de quantification afin d'atteindre l'objectif visé. Dans la première phase, on peut également constater un phénomène étonnant, qui est que l'on arrive à diminuer la distorsion tout en diminuant le débit. Ceci est dû au fait que partant d'une sur-segmentation initiale, il est très difficile d'obtenir un mouvement fiable pour les petites régions faiblement texturées; d'autre part, il n'est pas non plus évident que le meilleur mouvement au sens de la prédiction soit le meilleur mouvement au sens du codage. La phase de remise en cause du mouvement par l'intermédiaire du critère débit-distorsion permet alors de trouver des mouvements plus favorables.

3 Résultats

Nous avons testé notre schéma de codage orienté-régions sur différentes séquences d'images couleurs au format 4:2:0. Le pas de quantification des coefficients DCT utilisé dans la phase de codage des erreurs résiduelles a été initialisé à 10 au début de chaque optimisation (excepté pour la séquence *Mother&Daughter* où il a été initialisé à 4 du fait des faibles erreurs présentes). La mesure de distorsion que nous avons utilisé prend en compte les distorsions introduites sur les chrominances et est définie par: $D = EQM_Y + \frac{EQM_U + EQM_V}{4}$; les erreurs sur les chrominances sont divisées par 4 afin de prendre en compte leur sous-échantillonnage.

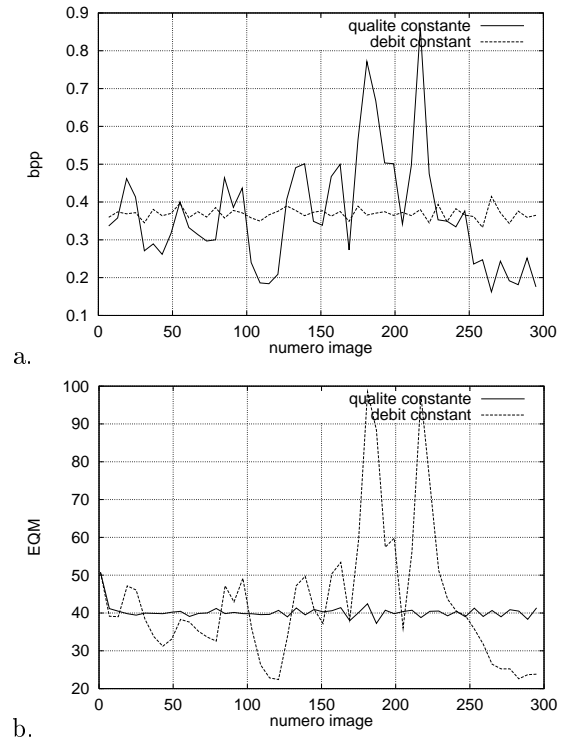


FIG. 2: Variations de débit (a) et distorsion (b) pour la séquence Foreman codée à 5 Hz; le débit visé est 0.37 bpp et la distorsion 40 (la distorsion est définie par $D = EQM_Y + \frac{EQM_U + EQM_V}{4}$).

La figure 2.a,b présente les variations en débit et distorsion obtenues pour le codage de la séquence *Foreman* codée à 5Hz, dans le cadre de codages à qualité constante et à débit constant. Les traitements pour chaque cas ont été réglés afin d'obtenir un débit moyen de 0.37 bpp ; on peut ainsi observer que le codage à qualité constante est en moyenne plus performant que le codage à débit constant (EQM respectives : 40 et 43.9, soit un écart de 0.4 dB). Comparé à un schéma de codage H263, on peut également observer pour cette gamme de débit, un gain de l'ordre de 1 dB sur la composante de luminance. La figure 5 présente la segmentation finale obtenue pour l'image 96 de la séquence, ainsi que la segmentation spatiale initiale utilisée et l'image décodée obtenue. Comparé aux précédentes études optimisant la segmentation selon des critères de type débit-distorsion [5, 6], les résultats obtenus sont supérieurs bien que notre distorsion soit basée sur les trois composantes, et non uniquement sur la luminance. Cette supériorité est liée au fait que nous ne sommes pas contraints dans notre approche par la donnée d'une segmentation hiérarchique qu'il faut respecter (quadtree pour [5], arbre de partitions spatio-temporelles emboîtées pour [6]).

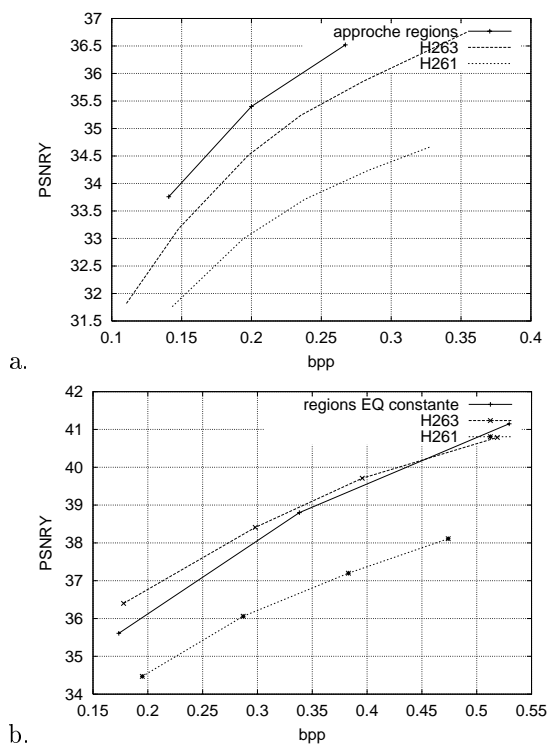


FIG. 3: Résultats obtenus sur les séquences *Suzie* (a) et *Mother&Daughter* (a) codée à 5 Hz (format QCIF 4:2:0). Le PSNR affiché est celui de la luminance, le débit est le débit total pour coder la séquence couleur.

La figure 3 présente les résultats obtenus sur les séquences *Suzie* et *Mother&Daughter* par notre schéma de codage et les schémas de codage H261 et H263. On peut ainsi voir que pour des séquences avec du mouvement (ie. *Suzie*, *Foreman*), notre schéma de codage permet d'obtenir un gain de l'ordre de 1 dB par rapport à H263 et de l'ordre de 2.5 dB par rapport à H261. Par contre, pour une sé-

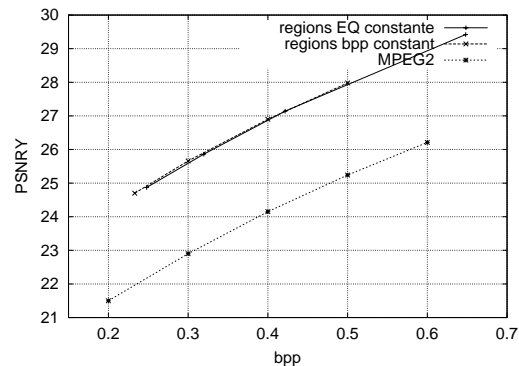


FIG. 4: Résultats obtenus pour le codage des 20 premières images de la séquence *Flower Garden* (format CIF 4:2:0). Le PSNR affiché est celui de la luminance, le débit est le débit total pour coder la séquence couleur.

quence avec un mouvement faible, tel *Mother&Daughter*, le résultat peut être légèrement inférieur à H263 pour les très bas débits, le gain par rapport à H261 restant toujours de l'ordre de 2 dB. Comparé à l'approche d'optimisation que nous avons proposée dans [4] basée sur le formalisme MDL (ie. minimisation du coût de description avec un pas de quantification fixé), pour un même débit final, et cette fois-ci pour des images couleurs, nous arrivons à une qualité équivalente pour la composante de luminance. Ceci aurait pu encore être amélioré, si l'on n'avait pris en compte que la distorsion de la composante de luminance, et non la distorsion sur les trois composantes.



FIG. 6: Comparaison entre l'approche proposée (a) et MPEG2 (b). Image 20 de la séquence *Flower Garden*, codée à 0.2 bpp.

Sur la figure 4, nous avons présenté une comparaison de



FIG. 5: Résultats obtenus sur l'image 96 de la séquence Foreman : segmentation spatiale initiale (a), segmentation finale obtenue après optimisation (b), image décodée (c).

notre schéma de codage avec MPEG dans le cadre d'une séquence télévisuelle (*Flower Garden*). On peut ainsi voir que le gain par rapport à MPEG2 est de presque 3 dB. La figure 6 présente l'image 20 décodée de la séquence *Flower Garden* obtenue par notre codeur et le codeur MPEG2. On peut ainsi voir que notre codeur permet d'obtenir plus de détails, et limite fortement les effets de blocs pour un même taux de compression.

4 Conclusion

Nous avons présenté dans cet article un nouveau schéma de codage orienté-régions optimisé selon un critère débit-distorsion. Les résultats expérimentaux obtenus sont très prometteurs, et montre que le schéma de codage proposé obtient des résultats supérieurs, voire équivalents dans le cas de faibles mouvements, au schéma de codage normalisé H263 pour des séquences visio-phoniques. Pour des séquences télévisuelles, les résultats obtenus sont nettement meilleurs que ceux obtenus par MPEG2 avec un gain pouvant aller jusqu'à 3 dB. Ainsi, bien qu'utilisant toujours une technique de codage de l'erreur résiduelle par blocs, les effets de blocs sont très nettement réduits par rapport aux schémas de codage MPEG et H26x pour un même débit.

L'optimisation du schéma de codage n'a été présentée dans ce papier que pour des images de type P, mais il serait également possible d'étendre cette technique à des images de type B, afin d'avoir une structure de GOP similaire à MPEG2.

Il est à noter tout de même que les gains obtenus dans cet article par rapport à des schémas de codage normalisés, sont à relativiser par la complexité algorithmique du codeur. Ainsi pour une séquence au format QCIF, il faut de l'ordre d'une vingtaine de minutes par images pour réaliser l'optimisation au codeur ; la complexité au décodeur quant à elle, reste quasiment identique à celle de MPEG ou H26x.

Références

[1] Bonnaud (L.), Pateux (S.) et Labit (C.). – Multiple objects tracking for efficient motion-based segmenta-

tion coding using a temporal prediction. In : *Proc. of PCS 97*, pp. 125–128. – Berlin, RFA, septembre 1997.

[2] Morier (F.), Benois-Pineau (J.), Barba (D.) et Sanson (H.). – Robust segmentation of moving image sequences. In : *ICIP'97*, pp. 719–722. – Santa Barbara, California, octobre 97.

[3] Moscheni (F.) et Bhattacharjee (S.). – Robust region merging for spatio-temporal segmentation. In : *ICIP'96*, pp. 501–504. – Lausanne, Switzerland, septembre 1996.

[4] Pateux (S.). – *Segmentation spatio-temporelle et codage orienté-régions de séquences vidéo basés sur le formalisme MDL*. – Thèse de doctorat, Université de Rennes I, 1998.

[5] Reusens (E.), Ebrahimi (T.), Le Buhan (C.), Castagno (R.), Vaerman (V.), Piron (L.), de Sola (C.), Bhattacharjee (S.), Bossen (F.) et Kunt (M.). – Dynamic approach to visual data compression. *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, n° 1, février 1997, pp. 197–211.

[6] Salembier (P.), Marqués (F.), Pardàs (M.), Morros (R.), Corset (I.), Jeannin (S.), Bouchard (L.), Meyer (F.) et Marcotegui (B.). – Segmentation-based video coding system allowing the manipulation of objects. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, n° 1, février 1997, pp. 60–73.

[7] Salembier (P.), Torres (L.), Meyer (F.) et Gu (C.). – Region-based video coding using mathematical morphology. In : *Proceedings of the IEEE*, éd. par IEEE. IEEE, pp. 843–857. – juin 1995. Special Issue on Digital Television.